Factor Analysis

Jacob Goldberger

Factor analysis model:

$$\begin{aligned} X &\sim N(0, I) \\ Y|X &\sim N(\Lambda X + \mu, \Psi) \end{aligned} \qquad \Psi \text{ is diagonal}$$

Statistical Inference:

$$x|y \sim N(\, V\Lambda^T \Psi^{-1}(y-\mu)\,,\, V\,) \qquad \text{such that} \qquad V = (I + \Lambda^T \Psi^{-1} \Lambda)^{-1}$$

Log Likelihood of $y_1, \ldots, y_N$:

$$\sum_t \log f(y_t|\Lambda, \Psi, \mu) = c - \frac{N}{2}|\Lambda\Lambda^T + \Psi| - \frac{1}{2}\sum_t (y_t - \mu)^T (\Lambda\Lambda^T + \Psi)^{-1}(y_t - \mu)$$

EM iteration:

- Expectation step :

$$\begin{aligned} \langle x_t \rangle &= E(x_t|y_t) \\ \langle x_t x_t^T \rangle &= E(x_t x_t^T|y_t) = V(x_t|y_t) + \langle x_t \rangle \langle x_t \rangle^T \end{aligned}$$

- Maximization step :

$$\begin{aligned} \hat{\Lambda} &= (\sum_t (y_t - \mu) \langle x_t \rangle^T)(\sum_t \langle x_t x_t^T \rangle)^{-1} \\ \hat{\Psi} &= diag \frac{1}{N} \sum_t (y_t - \mu - \hat{\Lambda} \langle x_t \rangle)(y_t - \mu)^T \end{aligned}$$

matrix identities: $A_{n\times n}, C_{k\times k}$ invertible, $B_{n\times k}, D_{k\times n}$

$$\begin{aligned} (A+BCD)^{-1} &= A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\ (A+BCD)^{-1}BC &= A^{-1}B(C^{-1} + DA^{-1}B)^{-1} \\ |A+BCD| &= |A||C||C^{-1} + DA^{-1}B| \end{aligned}$$

Let the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ be normally distributed according to:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} A & C^T \\ C & B \end{bmatrix} \right)$$

Then the marginal distributions are:

$$x \sim N(a, A) \qquad\qquad y \sim N(b, B)$$

and the conditional distributions are:

$$\begin{aligned} x|y &\sim N(a + C^T B^{-1}(y-b), A - C^T B^{-1} C) \\ y|x &\sim N(b + CA^{-1}(x-a), B - CA^{-1}C^T) \end{aligned}$$